# What is Amazon EC2 Auto Scaling?

Amazon EC2 Auto Scaling helps you ensure that you have the correct number of Amazon EC2 instances available to handle the load for your application. You create collections of EC2 instances, called Auto Scaling groups. You can specify the minimum number of instances in each Auto Scaling group, and Amazon EC2 Auto Scaling ensures that your group never goes below this size. You can specify the maximum number of instances in each Auto Scaling group, and Amazon EC2 Auto Scaling ensures that your group never goes above this size. If you specify the desired capacity, either when you create the group or at any time thereafter, Amazon EC2 Auto Scaling ensures that your group has this many instances. If you specify scaling policies, then Amazon EC2 Auto Scaling can launch or terminate instances as demand on your application increases or decreases.

For example, the following Auto Scaling group has a minimum size of one instance, a desired capacity of two instances, and a maximum size of four instances. The scaling policies that you define adjust the number of instances, within your minimum and maximum number of instances, based on the criteria that you specify.


   An illustration of a basic Auto Scaling group.
  

## Auto Scaling components

The following table describes the key components of Amazon EC2 Auto Scaling.

|  |  |
| --- | --- |
| A graphic representing an Auto Scaling group. | **Groups**  Your EC2 instances are organized into groups so that they can be treated as a logical unit for the purposes of scaling and management. When you create a group, you can specify its minimum, maximum, and, desired number of EC2 instances. For more information, see [Auto Scaling groups](https://docs.aws.amazon.com/autoscaling/ec2/userguide/AutoScalingGroup.html). |
| A graphic representing a launch configuration. | **Configuration templates**  Your group uses a launch template, or a launch configuration (not recommended, offers fewer features), as a configuration template for its EC2 instances. You can specify information such as the AMI ID, instance type, key pair, security groups, and block device mapping for your instances. For more information, see [Launch templates](https://docs.aws.amazon.com/autoscaling/ec2/userguide/LaunchTemplates.html) and [Launch configurations](https://docs.aws.amazon.com/autoscaling/ec2/userguide/LaunchConfiguration.html). |
| A graphic representing scaling options. | **Scaling options**  Amazon EC2 Auto Scaling provides several ways for you to scale your Auto Scaling groups. For example, you can configure a group to scale based on the occurrence of specified conditions (dynamic scaling) or on a schedule. For more information, see [Scaling options](https://docs.aws.amazon.com/autoscaling/ec2/userguide/scaling_plan.html#scaling_typesof). |

## Getting started

If you're new to Amazon EC2 Auto Scaling, we recommend that you review [Amazon EC2 Auto Scaling instance lifecycle](https://docs.aws.amazon.com/autoscaling/ec2/userguide/AutoScalingGroupLifecycle.html) before you begin.

To begin, complete the [Getting started with Amazon EC2 Auto Scaling](https://docs.aws.amazon.com/autoscaling/ec2/userguide/GettingStartedTutorial.html) tutorial to create an Auto Scaling group and see how it responds when an instance in that group terminates. If you already have running EC2 instances, you can create an Auto Scaling group using an existing EC2 instance, and remove the instance from the group at any time.

## Accessing Amazon EC2 Auto Scaling

If you've signed up for an Amazon Web Services account, you can access Amazon EC2 Auto Scaling by signing into the AWS Management Console, choosing **EC2** from the console home page, and then choosing **Auto Scaling Groups** from the navigation pane.

You can also access Amazon EC2 Auto Scaling using the [Amazon EC2 Auto Scaling API](https://docs.aws.amazon.com/autoscaling/ec2/APIReference/). Amazon EC2 Auto Scaling provides a Query API. These requests are HTTP or HTTPS requests that use the HTTP verbs GET or POST and a Query parameter named Action. For more information about the API actions for Amazon EC2 Auto Scaling, see [Actions](https://docs.aws.amazon.com/autoscaling/ec2/APIReference/API_Operations.html) in the Amazon EC2 Auto Scaling API Reference.

If you prefer to build applications using language-specific APIs instead of submitting a request over HTTP or HTTPS, AWS provides libraries, sample code, tutorials, and other resources for software developers. These libraries provide basic functions that automate tasks such as cryptographically signing your requests, retrying requests, and handling error responses, making it is easier for you to get started. For more information, see [AWS SDKs and tools](https://aws.amazon.com/tools/).

If you prefer to use a command line interface, you have the following options:

**AWS Command Line Interface (CLI)**

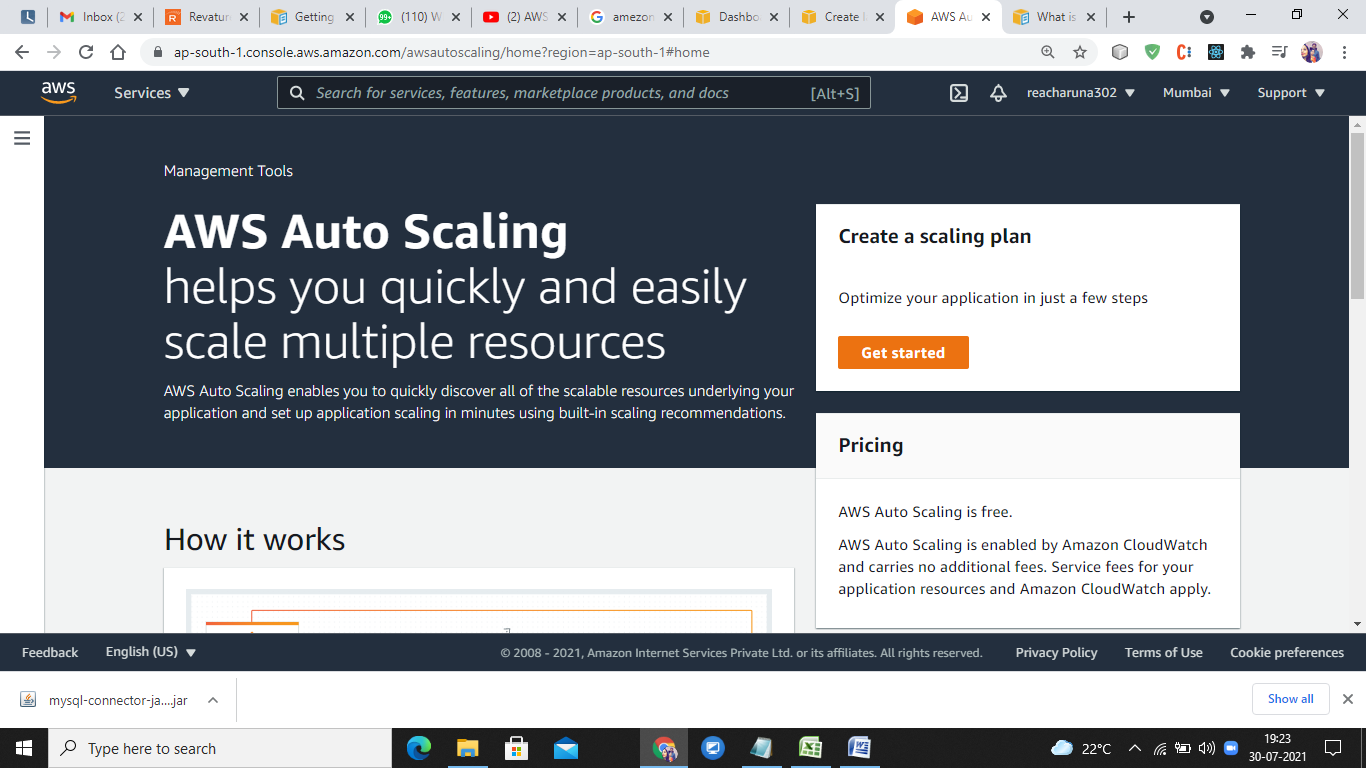
Provides commands for a broad set of AWS products, and is supported on Windows, macOS, and Linux. To get started, see [AWS Command Line Interface User Guide](https://docs.aws.amazon.com/cli/latest/userguide/). For more information, see [autoscaling](https://docs.aws.amazon.com/cli/latest/reference/autoscaling/index.html) in the AWS CLI Command Reference.

**AWS Tools for Windows PowerShell**

Provides commands for a broad set of AWS products for those who script in the PowerShell environment. To get started, see the [AWS Tools for Windows PowerShell User Guide](https://docs.aws.amazon.com/powershell/latest/userguide/). For more information, see the [AWS Tools for PowerShell Cmdlet Reference](https://docs.aws.amazon.com/powershell/latest/reference/Index.html).

For information about your credentials for accessing AWS, see [AWS security credentials](https://docs.aws.amazon.com/general/latest/gr/aws-security-credentials.html) in the Amazon Web Services General Reference. For information about regions and endpoints for calls to Amazon EC2 Auto Scaling, see the [Regions and endpoints](https://docs.aws.amazon.com/general/latest/gr/as.html) table in the AWS General Reference.

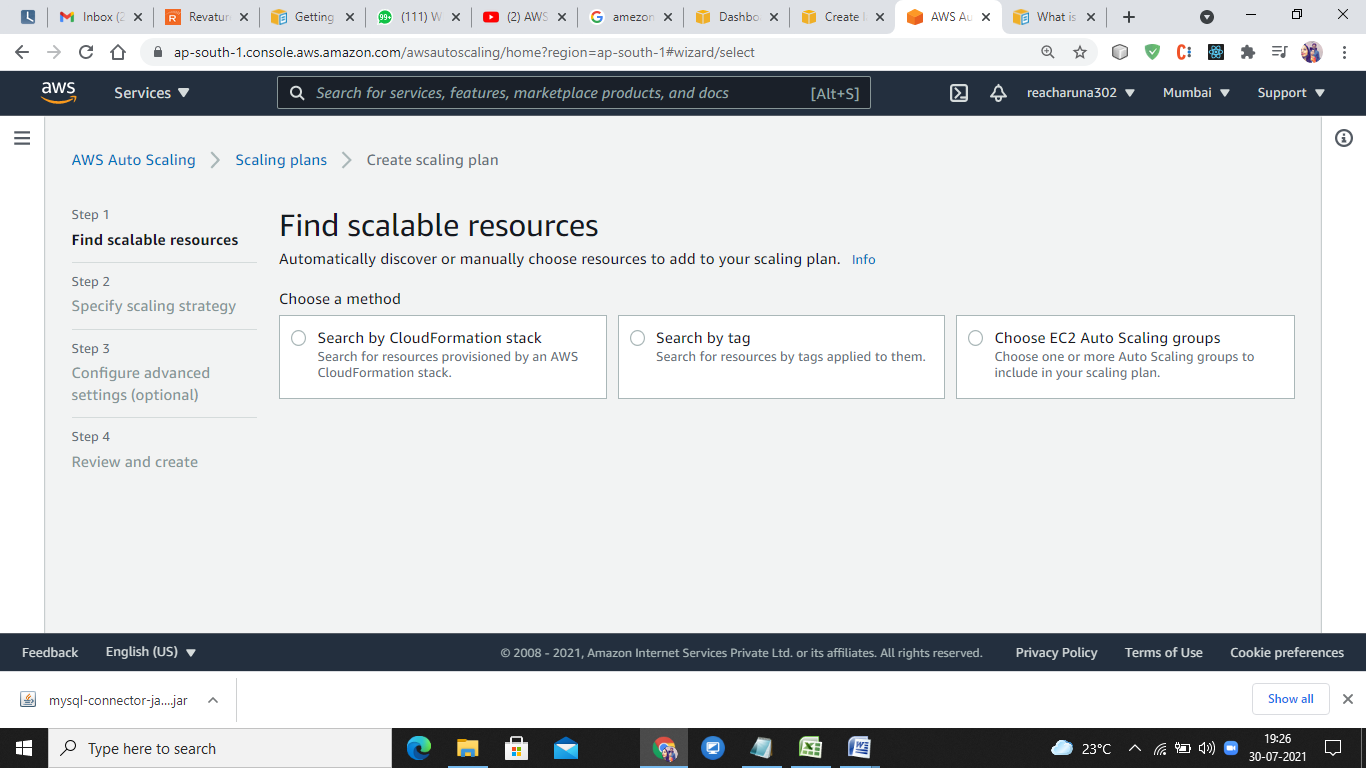
Search for a service AWS Auto Scaling

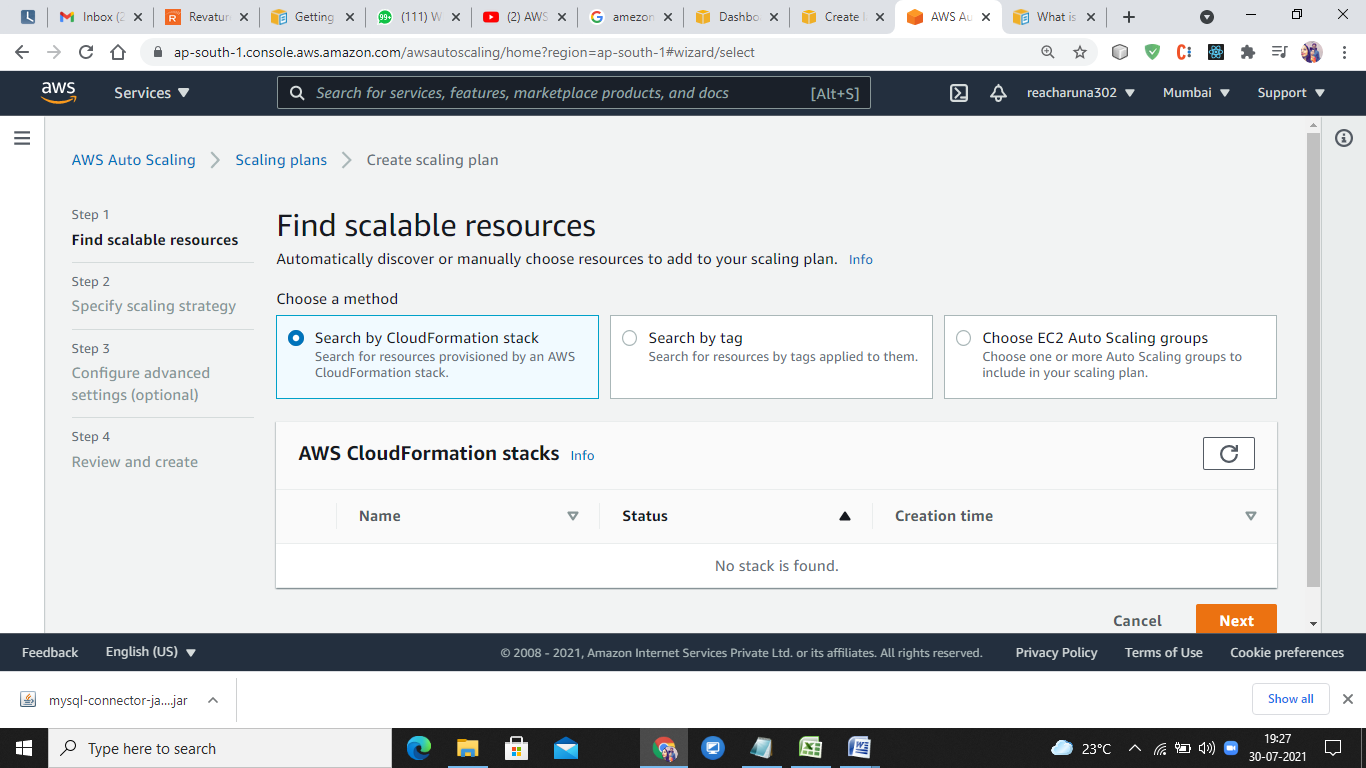


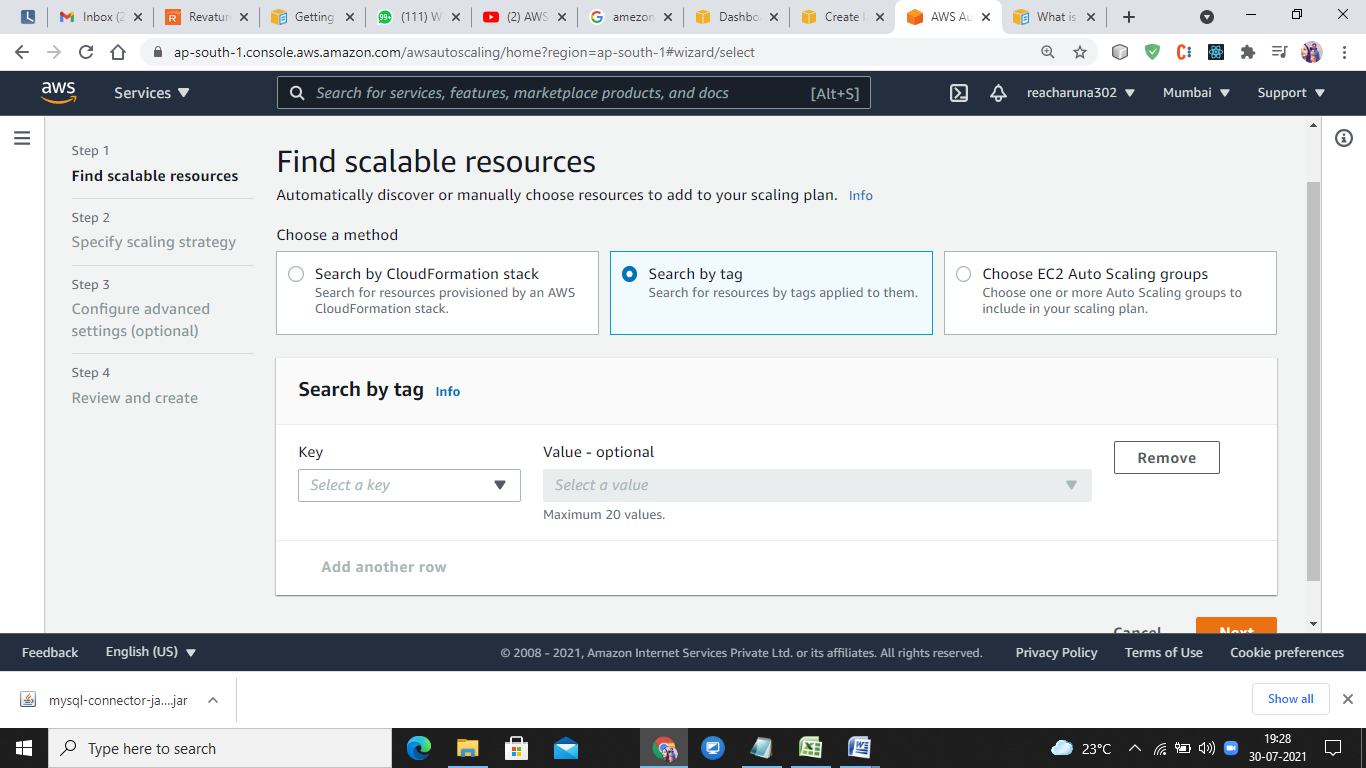
Step->

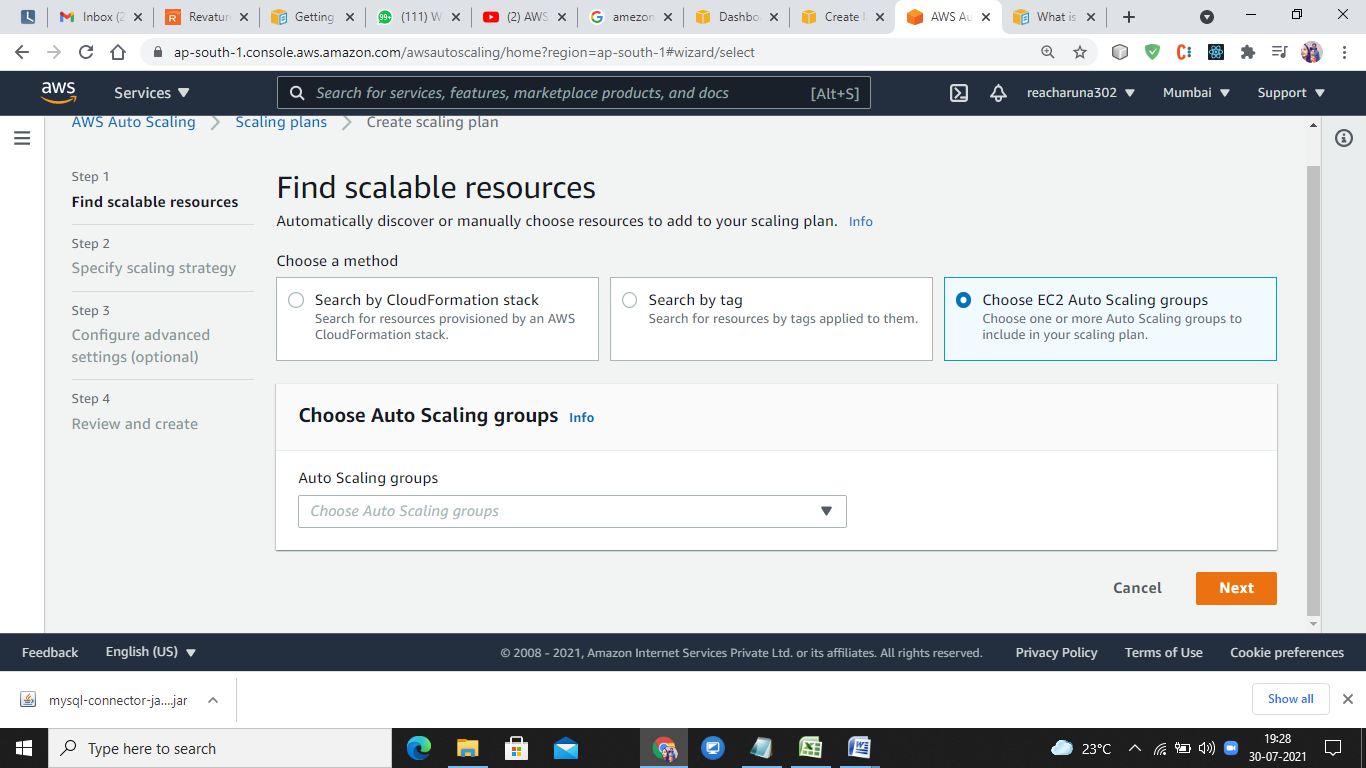
Find scalable resources

Automatically discover or manually choose resources to add to your scaling plan









AWS Auto Scaling

With AWS Auto Scaling, you configure and manage scaling for your scalable AWS resources through a scaling plan. It lets you choose scaling strategies to define how to optimize your resource utilization. You can optimize for availability, for cost, or a balance of both. Alternatively, you can leverage custom strategies for greater

Benefits

Maintain performance

Right-size multiple resources to meet application performance targets.

Control costs

Automatically remove unneeded resources to avoid overspending.

Simplify scaling

Discover scalable resources and build a scaling plan in just a few steps.

Scale predictively

Use historical load metrics to learn your application’s traffic patterns and proactively scale resources ahead of time. Currently only available for EC2 Auto Scaling groups.

## Pricing

AWS Auto Scaling is free.

AWS Auto Scaling is enabled by Amazon CloudWatch and carries no additional fees. Service fees for your application resources and Amazon CloudWatch apply